Analysis on the timbre of Ambisonics recording by circular and spherical microphone array using a binaural loudness model

Bosun XIE[1]; Yang LIU[2]

Acoustic Lab., Physics Dept., South China University of Technology, Guangzhou, 510641, China

ABSTRACT

Ambisonics are a series of flexible sound reproduction systems that decompose and reconstruct sound field by each order approximation of spatial harmonics decomposition. Circular or spherical microphone array has been used for Ambisonics recording. For an array with given radius as well as given number and configuration of microphones, the spatial harmonics signals up to certain temporal frequency and certain order can be derived from the array outputs with appropriate accuracy. When the temporal frequency exceeds the given limitation of array, however, spatial aliasing occurs in the derived spatial harmonics signals. Spatial aliasing in recording causes error in reconstructed sound field, which inevitably change the timbre in reproduction. The present work analyzes the timbre in Ambisonics recording by rigid circular and spherical microphone array. The binaural loudness level spectra are used as a criterion to evaluate the timbre in reproduction. The results indicate that, array recording influences little on the binaural loudness level spectra and thus timbre in final reconstructed sound field providing that the high-frequency limit of microphone array exceeds that of sound field reconstruction. Increasing number of microphones extends the high-frequency limit of array and thus improves the timbre.

Keywords: Ambisonics, Timbre, Microphone array   I-INCE Classification of Subjects Number(s): 61.2

1. INTRODUCTION

Ambisonics are a series of flexible sound reproduction systems that use spatial harmonics decompositions of sound field as independent signals (1). The independent signals are decoded (linearly combined) and then reproduced by circular or spherical arrangement of loudspeakers. For a given order, Ambisonics are able to reconstruct target sound field within a region around the origin and below certain temporal frequency limit. Beyond the frequency limit, spectral error in reproduction stage occurs.

Circular or spherical microphone array has been used for recording the spatial harmonics signals (2, 3). Given the array radius as well as the number and configuration of microphones, the spatial harmonics signals up to certain order and certain temporal frequency can be derived from the array outputs with appropriate accuracy. When the temporal frequency exceeds the given high–frequency limit of the array, however, spatial aliasing occurs in the derived spatial harmonics signals.

Both spatial aliasing in recording and spectral error in reproduction cause errors in final reconstructed sound field, thus inevitably change the timbre in reproduction. Some previous works have analyzed the physical errors in microphone array recording and Ambisonics reproduction. In present work, we use Moore's modified loudness model to evaluate the timbre coloration in Ambisonics, especially those caused by circular and spherical microphone array recording.

2. AMBISONICS AND MICROPHONE ARRAY RECORDING

A counterclockwise spherical coordinate system is used in this paper. The origin of coordinates is located at the centre of the head. The sound source position is specified by distance $0 \leq r \leq +\infty$, elevation $-90° \leq \phi \leq 90°$ and azimuth $0° \leq \theta < 360°$. Where $\phi = 0°$ and $90°$, respectively, correspond to the horizontal plane and top direction. In the horizontal plane, $\theta = 0°$ is the front and $\theta = 90°$ is the left.

---

[1] phbsxie@scut.edu.cn

[2] shenhua.liuyang@163.com

The target sound field for a single incident plane wave is discussed because any sound field in source-free region can be decomposed into superposition of plane waves from different incident directions. For the Q order horizontal Ambisonics (4), suppose that M loudspeakers are arranged uniformly in a circle around the origin. Spatial sampling theorem requires $M \geq (2Q+1)$. The loudspeaker signals are expressed as a Fourier series of azimuth θ_S of incident plane wave and then truncated up to order Q. In the case of far-field loudspeaker layout, the signal for the ith loudspeaker at azimuth θ_i is given by:

$$E_i(\theta_S) = \frac{1}{M}[S_0^{(1)} + 2\sum_{q=1}^{Q}[\cos q\theta_i S_q^{(1)} + \sin q\theta_i S_q^{(2)}] \quad (1)$$

and

$$S_0^{(1)} = E_0 \quad S_q^{(1)} = E_0 \cos q\theta_S \quad S_q^{(2)} = E_0 \sin q\theta_S \quad (2)$$

Where E_0 is a constant related to total signal amplitude. Eq. (2) represents set of $(2Q+1)$ independent signals with various orders of directivity.

Similarly, the loudspeaker signals for $(L-1)$ order spatial Ambisonics can be expressed as a linear combinations of L^2 spherical harmonic functions of incident direction up to order $(L-1)$ (4). When M loudspeakers are arranged on a spherical surface around the listener, the minimum number M depends on the order $(L-1)$ and loudspeaker layout, but at least should fulfill $M \geq L^2$. Let $\Omega_S = (\theta_S, \phi_S)$ denote the target direction of incident plane wave, and $\Omega_i = (\theta_i, \phi_i)$ denote the direction of ith loudspeaker. In the case of far-field and nearly uniform loudspeaker layout, the signal of ith loudspeaker is given by:

$$E_i(\Omega_S) = \frac{4\pi}{M}\sum_{l=0}^{L-1}\sum_{m=0}^{+l}\sum_{\sigma=1}^{2} Y_{lm}^{(\sigma)}(\Omega_i) S_{lm}^{(\sigma)}(\Omega_S) \quad (3)$$

Where $S_{lm}^{(\sigma)}(\Omega_S)$ is a set of L^2 independent signals with various orders of directivity,

$$S_{lm}^{(\sigma)} = E_0 Y_{lm}^{(\sigma)}(\Omega_S) \quad (4)$$

And $Y_{lm}^{(\sigma)}(\Omega)$ are real-valued spherical harmonics functions of order l and degree m at target direction Ω_S,

$$Y_{lm}^{(\sigma)}(\Omega_S) = N_{lm} P_l^{|m|}[\cos(90^\circ - \phi)] \begin{cases} \cos m\theta & \sigma = 1 \\ \sin m\theta & \sigma = 2 \end{cases}$$

$$N_{lm} = \sqrt{\frac{(l-|m|)!(2l+1)}{(l+|m|)!2\pi\Delta_m}} \qquad \Delta_M = \begin{cases} 2 & m = 0 \\ 1 & m \neq 0 \end{cases} \quad (5)$$

$P_l^{|m|}[\cos(90^\circ - \phi)]$ is the associated Legendre polynomials.

The independent signals in Eq. (2) and Eq. (4) can be recorded by a circular or spherical microphone array. In the case of horizontal Ambisonics, M' omnidirectional microphones are arranged uniformly in a horizontal circle on a rigid cylinder surface with infinite length and radius r_M. Let $E_{i'}(\theta_{i'}, kr_M, \theta_S)$ denote the output of i' th microphone at azimuth $\theta_{i'}$. Then the independent signals in Eq. (2) can be derived from a weighted combination of M' microphone outputs (2):

$$S_q^{(\sigma)} = \frac{1}{M' R_q(kr_M)}\sum_{i'=0}^{M'-1} E_{i'}(\theta_{i'}, kr_M, \theta_S) \begin{cases} \cos q\theta_{i'} & \sigma = 1 \\ \sin q\theta_{i'} & \sigma = 2 \end{cases} \quad (6)$$

Where

$$R_q(kr_M) = j^q \left[J_q(kr_M) - \frac{dJ_q(kr_M)/d(kr_M)}{dH_q(kr_M)/d(kr_M)} H_q(kr_M) \right] \quad (7)$$

$J_q(kr_M)$ and $H_q(kr_M)$ are q-order Bessel function and Hankel function of the second kind, respectively; k is the wave number.

Similarly, in the case of spatial Ambisonics, M' omnidirectional microphones are arranged on a rigid spherical surface with radius r_M. Let $E_{i'}(\Omega_{i'}, kr_M, \Omega_S)$ denote the output of i' th microphone at direction $\Omega_{i'}$. For nearly uniform microphone layout, the independent signals in Eq. (4) can be derived from a weighted combination of M' microphone outputs (3):

$$S_{lm}^{(\sigma)}(\Omega_S) = \frac{4\pi}{MR_l(kr_M)} \sum_{i'=0}^{M'-1} E'_{i'}(kr_M, \Omega_{i'}, \Omega_S) Y_{lm}^{(\sigma)}(\Omega_{i'})$$
$$l = 0,1,2 \ldots (L-1) \quad m = 0,1..l \qquad \sigma = 1,2 \tag{8}$$

Where

$$R_l(kr_M) = 4\pi\, j^l \left[j_l(kr_M) - \frac{dj_l(kr_M)/d(kr_M)}{dh_l(kr_M)/d(kr_M)} h_l(kr_M) \right] \tag{9}$$

$j_l(kr_M)$ and $h_l(kr_M)$ are l-order spherical Bessel function and spherical Hankel function of the second kind, respectively.

Suppose that the independent signals in Eq. (2) and Eq. (4) are accurate. Q order horizontal Ambisonics or (L-1) order spatial Ambosonics is able to reconstruct target sound field with reasonable accuracy below a high-frequency limit $f_{max,H}$ and within a circular or spherical region with radius r_H:

$$f < f_{max,H} = \frac{Qc}{2\pi r_H} \quad (horizontal) \quad or \quad f < f_{max,H} = \frac{(L-1)c}{2\pi r_H} \quad (spatial) \tag{10}$$

Where c is the speed of sound. It can be seen that the product of $f_{max,H}$ and r_H increases with the order of Ambisonics.

However, for a rigid circular or spherical microphone array with radius r_M and an uniform layout of M' microphone, the independent signals derived from Eq. (6) and Eq. (8) are accurate only below a high-frequency limit $f_{max,M}$ that is imposed by spatial sampling theorem,

$$f < f_{max,M} = \frac{M'c}{4\pi r_M} \quad (horizontal) \quad or \quad f < f_{\max,M} = \frac{(\sqrt{M'}-1)c}{2\pi r_M} \quad (spatial) \tag{11}$$

Otherwise, spatial aliasing error occurs in the independent signals derived from Eq. (6) and Eq. (8).

Therefore, the final high-frequency limit for accurate sound field reconstruction depends on these of both microphone array in recording and Ambisonics reproduction. When the temporal frequency of target plane wave exceeds either the high-frequency limit of microphone array or that of sound field reproduction,

$$f > f_{\max} = \min(f_{max,M}, f_{max,H}) \tag{12}$$

error occurs the final reconstructed sound field which inevitably results in perceivable timbre coloration.

3. ANALYSIS OF THE TIMBRE COLORATION IN THE FINAL RECONSTRCTED SOUND FIELD

Moore's modified loudness model (5), in which the inhibition between the binaural sound signals has been considered, is used to analyze the binaural loudness level spectra and then timbre coloration in Ambisonics recording and reproduction. The details of using Moore's modified loudness model to analyze the timbre coloration are referred to (6). The procedure of the analysis in this paper includes following steps:

- Calculate the binaural pressures in final reconstructed sound field.
- Calculate the binaural loudness level spectra from binaural pressures using Moore's modified loudness model.
- Compare the binaural loudness level spectra for final reconstructed sound field with those of the target sound field.

For a central listening position and incident plane wave from direction (θ_S, ϕ_S), the binaural pressures can be evaluated using the far-field head-related transfer functions (HRTFs) (7):

$$P_L(f) = HRTF_L(\theta_s, \phi_s, f) S_0(f) \quad P_R(f) = HRTF_R(\theta_s, \phi_s, f) S_0(f) \tag{13}$$

Where $S_0(f)$ is the input stimulus.

The summing binaural pressures for final reconstructed sound field can be evaluated by:

$$P'_L(f) = \sum_{i=1}^{M} HRTF_L(\theta_i, \phi_i, f) E_i(\theta_S, \phi_S)$$
$$P'_R(f) = \sum_{i=1}^{M} HRTF_R(\theta_i, \phi_i, f) E_i(\theta_S, \phi_S) \quad (14)$$

Where $HRTF_L(\theta_i, \phi_i, f)$ and $HRTF_R(\theta_i, \phi_i, f)$ are HRTFs from the ith loudspeaker at far field distance to the left and right ear, respectively.

Then the binaural loudness level spectra in each ERB (equivalent rectangular bandwidth) of auditory filter) is calculated. The relationship between the ERB and the center frequency of auditory filter (in kHz) is (8):

$$ERB = 24.7(4.37f + 1) \quad (15)$$

In view of the equivalent rectangular auditory filters, a new frequency scale called ERB Number can be introduced. The relationship between ERB Number and frequency f (in kHz) is:

$$ERBN = 21.4 \log_{10}(4.37f + 1) \quad (16)$$

The binaural loudness level spectra describe the perceived loudness level in different frequency bands. When the difference between the binaural loudness level spectra in final reconstructed sound field and that in target field exceeds 1 Phon/ERB (Just noticeable difference, JND), the timbre coloration can be perceived (9). This criterion is used in analysis the timbre coloration in microphone array recording and Ambisonics reproduction.

4. RESULT AND DISCUSSTION

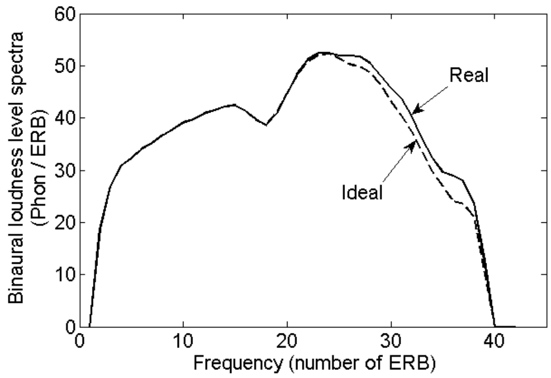
A mono pink noise was used as the stimulus and scaled to be equivalent to a free-field pressure level of 70 dB. The distances from loudspeakers to the head center are all 1.45 m. The HRTFs of KEMAR artificial head that were calculated using boundary element method with an azimuthal resolution of 1 °were adopted (10).

4.1 Horizontal Ambisonics

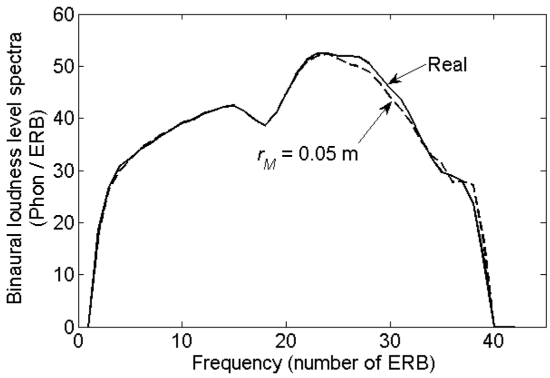
The Q = 5 order horizontal Ambisonics is taken as an example. The independent signals are recorded using a circular array with 12 microphones and various radius r_M. Then the decoded signals are reproduced by 24 loudspeakers. The target sound field is incident plane wave at azimuth θ_S = 45 ° in the horizontal plane. Fig. 1 (a) to (e) shows the binaural loudness level spectra (BLLS) of Ambisonics reproduction using the ideal signals given by Eq. (2) and these recorded by microphone array with radius r_M = 0.05 m, 0.1 m, 0.2 m and 0.3 m respectively. For comparison, the results of a real incident plane wave from θ_S = 45 °are also plotted in the figures. In the figure, the abscissa represents frequency in number of ERB, and the ordinate is the BLLS in Phon/ERB.

It can be observed from Fig. 1(a) to (e) that the BLLS in final reconstructed sound field of Ambisonics deviates from this of real plane wave above some frequency. The second row of Table 1 lists the high-frequency limit at which the deviation of BLLS exceeds the JND of 1 Phon/ERB. The high-frequency limits for independent signals recorded by microphone array with radius 0.05 m or 0.1 m are similar to this of ideal case. When the array radius exceeds 0.1 m, however, the high-frequency limit steadily decreases with increase of array radius. For target sound field of incident plane wave at other directions, the results are similar to above.

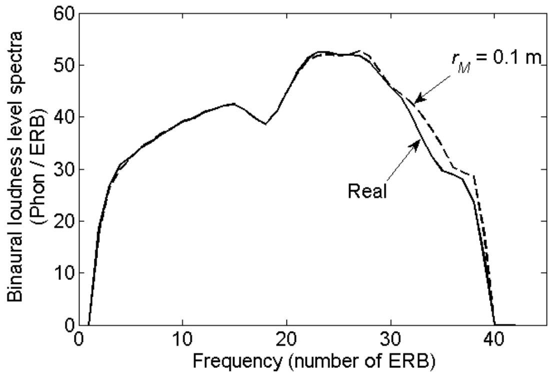
Consider a circular region with radius r_H = 0.0875 m that is matched to the average size of human head. According to Eq. (10), the high-frequency limit of the reproduction stage is $f_{max,H}$ = 3.16 kHz for Q = 5 order Ambisonics. Moreover, for an array with various radius and M' = 12 microphones, the high-frequency limit of the array is calculated by Eq. (11). Then the high-frequency limit of final reconstructed sound field is evaluated by Eq. (12). The results are a little bit less than these evaluated from BLLS calculation, as listed in the third row of Table 1. These results are reasonable because BLLS analysis yields the just perceivable error while Eq. (12) yields the high-frequency limit of accurate reconstructed sound field.



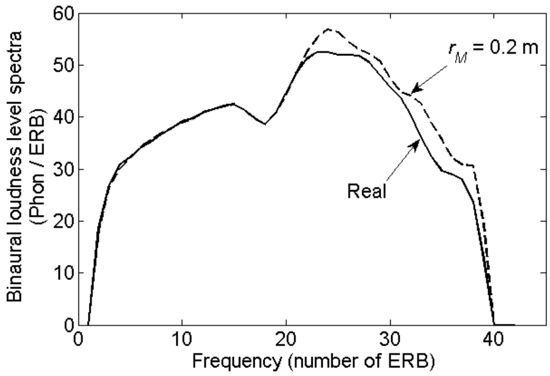
(a) Ideal



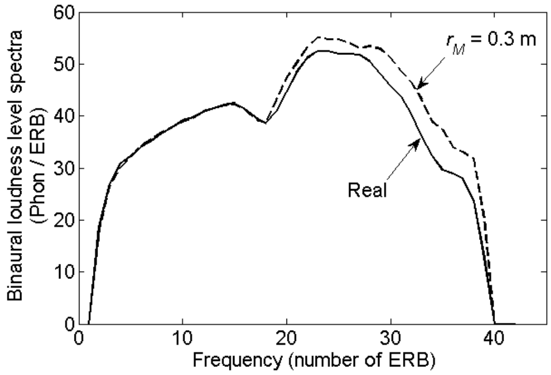
(b) r_M = 0.05 m



(c) r_M =0.1 m



(d) r_M = 0.2 m



(e) r_M =0.3 m

Figure 1 – BLLS of Ambisonics reproduction at θ_S = 45 °

Table 1 – The high-frequency limit for perceived timbre

Independent signals	Ideal	From array with 12 microphones and various radius				From array with 24 microphones and radius 0.2 m
		0.05 m	0.1 m	0.2 m	0.3m	
High-frequency limit evaluated from BLLS in ERBN (or kHz)	25 (3.14)	25 (3.14)	26 (3.53)	21 (1.96)	18 (1.36)	25 (3.14)
f_{max} in kHz evaluated from Eq. (12)	3.12	3.12	3.12	1.64	1.09	

Actually, providing that the high-frequency limit of microphone array is higher than this of Ambisonics reproduction, below the high-frequency limit of Ambisonics reproduction, the influence of spatial aliasing error of microphone array on the final reconstructed sound field is neglectable. Eq. (10) and Eq. (11) and the condition of $f_{max,M} > f_{max,H}$ yield:

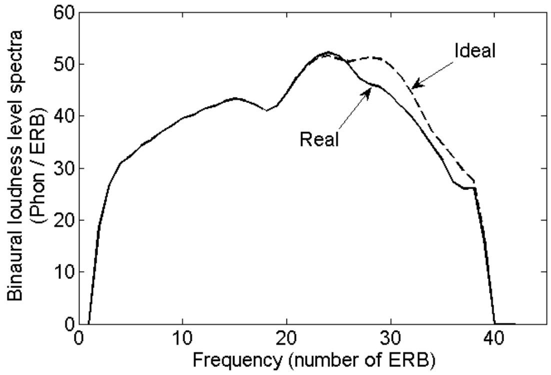
$$\frac{M'}{r_M} > \frac{2Q}{r_H} \tag{17}$$

For Q = 5, M' =12 and r_H = 0.0875 m, Eq. (17) yileds r_M < 0.105 m which is consistent with the results from BLLS analysis.

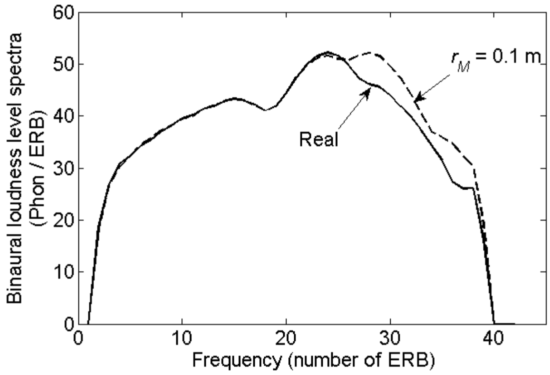
Eq. (17) is applicable to the design of microphone array for Ambisonics recording. It has been proved that a microphone array with large radius is preferred in view of improving the low-frequency signal-noise ratio in higher-order Ambisonics recording. However, Eq. (17) implies that a microphone array with large radius requires more microphones to avoid spatial aliasing in recording. For example, for an array with 24 microphones and radius 0.2 m, the last column of Table 1 lists the high-frequency limit evaluated from BLLS analysis and Eq. (12). Therefore, increasing array radius and number of microphone in array improve the low-frequency performance of signals, and at the same time, reduces the change in binaural loudness level spectra at high frequency. In other words, increasing array radius and number of microphone in array improves the timbre in reproduction. Of course, a tradeoff between the timbre performance in reproduction and the complexity of the microphone array should be made.

4.2 Spatial Ambisonics

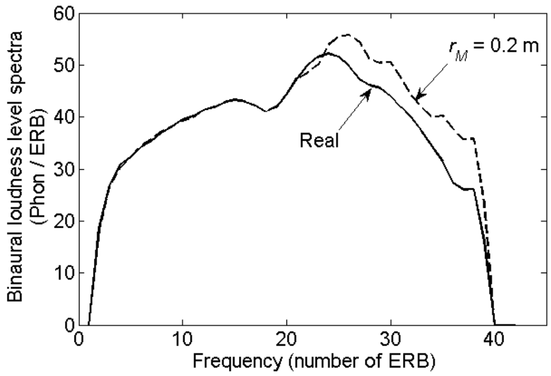
The Q = 6 order spatial Ambisonics is taken as an example. The independent signals are recorded using a spherical microphone array with 64 microphones and various radius r_M. Then the decoded signals are reproduced by 128 loudspeakers with nearly uniform layout. The target sound field is incident plane wave at (θ,ϕ) = (90°, 90°). Fig. 2 (a) to (c) shows the BLLS of Ambisonics reproduction using the ideal signals given by Eq. (3) and these recorded by a spherical microphone array with radius r_M = 0.1 m and 0.2 m, respectively. For comparison, the results of a real incident plane wave from (θ,ϕ) = (90°, 90°) are also plotted in the figures.



(a) Ideal



(b) r_M = 0.1 m



(c) r_M = 0.2 m

Figure 2 – BLLS of Ambisonics reproduction at (θ,ϕ) = (90°, 90°)

Similar to the case of horizontal Ambisonics, the BLLS in final reconstructed sound field of spatial Ambisonics deviates from this of real plane wave above some frequency. The high-frequency limit at which the deviation of BLLS exceeds the JND of 1 Phon/ERB are 26, 25 and 21 ERBN, or equally, 3.53, 3.14 and 1.96 kHz, for ideal signals and signals recorded by a spherical microphone array with radius r_M = 0.1 m and 0.2 m, respectively. When the array radius exceeds 0.1 m, the high-frequency limit steadily decreases with increase of array radius. For target sound field of incident plane wave at other directions, the results are similar to above. The physical reason and analysis of the results are similar to the case of horizontal Ambisonics and are omitted here.

5. CONCLUSIONS

Ambisonics utilize spatial harmonics of sound field as independent signals. Circle or spherical microphone array has been used for Ambisonics recording. The accuracy and then timbre in the final reproduced sound field depend on both the accuracy in the stage of sound field reconstruction and that in the stage of recording. For a given reproduction order and region (such as head radius), providing that the high-frequency limit of microphone array exceeds that of sound field reconstruction, array recording influences little on the binaural loudness level spectra and thus timbre in final reproduction. Otherwise, timbre coloration in final reproduction increases. Because the high-frequency limit of microphone array increases with the number of microphone, increasing number of microphone in array reduces the change in binaural loudness level spectra at high frequency and improves the timbre.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (11174087), and Guangzhou science and technology plan projects (No. 2011DH014). Authors also thank Prof. Xiaoli Zhong for her help.

REFERENCES

1. Gerzon MA. Ambisonics in multichannel broadcasting and video. J.Audio Eng.Soc. 1985; 33(11): 859-871.
2. Poletti M. A unified theory of horizontal holographic sound systems. J.Audio Eng.Soc. 2000; 48(12): 1155-1182.
3. Rafaely B. Plane-wave decomposition of the sound field on a spherical by convolution. J.Acoust.Soc.Am. 2004; 116(4): 2149-2157.
4. Moreau S, Daniel J, Bertet S. 3D sound field recording with higher order Ambisonics - objective measurements and validation of spherical microphone. Proc the AES 120th Convention; 20-23 May 2006; Paris, France, 2006. p. 6857.
5. Moore BCJ, Glasberg BR. “Modeling binaural loudness”. J.Acoust. Soc. Am. 2007; 121(3): 1604-1612.
6. Xie BS, Zhang TT. The audibility of spectral detail of head-related transfer functions at high frequency. Acta Acustica United with Acustica. 2010; 96(2): 328-339.
7. Xie BS. Head-related transfer function and virtual auditory display (second edition). USA: J Ross Publishing; 2013.
8. Moore BCJ. An introduction to the psychology of hearing (Sixth edition). UK: Emerald Group Publishing Limited; 2012.
9. Florentine M, Buus S, Mason CR. "Level discrimination as a function of level for tones from 0.25 to 16 kHz". J. Acoust. Soc. Am. 1987; 81(5): 1528-1541.
10. Liu Y, Xie BS., Yu GZ, et al. "Analysis on spatial discrimination threshold of head-related transfer function magnitude". To be published in Acta Acustica (in Chinese).